

EFFICIENT ROBUST ESTIMATION OF  
TIME-SERIES REGRESSION MODELS

PAVEL ČÍŽEK, Tilburg

(Contributed)

*Abstract.* The paper studies a new class of robust regression estimators based on the two-step least weighted squares (2S-LWS) estimator which employs data-adaptive weights determined from the empirical distribution or quantile functions of regression residuals obtained from an initial robust fit. Just like many existing two-step robust methods, the proposed 2S-LWS estimator preserves robust properties of the initial robust estimate. However, contrary to the existing methods, the first-order asymptotic behavior of 2S-LWS is fully independent of the initial estimate under mild conditions. We propose data-adaptive weighting schemes that perform well both in the cross-section and time-series data and prove the asymptotic normality and efficiency of the resulting procedure. A simulation study documents these theoretical properties in finite samples.

*Keywords:* asymptotic efficiency, least weighted squares, robust regression, time series

*MSC 2000:* 62F10, 62F12, 62F35, 62J05

## 1. INTRODUCTION

In statistics and econometrics, more and more attention is paid to techniques that can deal with data containing atypical observations, which can arise from outliers, miscoding, or heterogeneity not captured or presumed in a model. In the context of regression models, this is of high importance especially in time series as the least squares (LS) and maximum likelihood (MLE) estimators are heavily influenced by data contamination. For example, using real economic data, [1] documents presence of outliers in macroeconomic time series and [11], [13] evidence data contamination in financial time series and its adverse effects on estimators (e.g., quasi-maximum likelihood) and tests, respectively. On the other hand, the use of methods robust to atypical observations is infrequent (in econometrics) and usually limited to detection of outliers even in recent applications [12], [16] although exceptions exist [8]. The reasons could range from missing particular results regarding robust inference,

low relative efficiency of many robust methods, or the necessity to choose auxiliary tuning parameters without rigorous guidance. In addition, even the straightforward detection of outliers by a robust method or eye-balling and, after removing outliers, a subsequent application of a standard method such as least squares (as suggested in [10]) is not a theoretically justified inference method as the usual standard errors (and statistics based on them) will be biased. This strategy was however enhanced and thoroughly studied in [6], whose authors proposed and employed a data-dependent cut-off point for detecting outliers. Their robust and efficient weighted least squares (REWLS) estimator is asymptotically efficient if errors are normally distributed, but in a general case, the asymptotic distribution of REWLS depends on the initial estimator (in a known or unknown way, depending on the initial estimator).

To address these issues, we proposed in [4] a new class of robust estimation methods, the two-step least weighted squares (2S-LWS), which rely on an initial robust estimate and preserve its robust properties. Contrary to the existing methods, 2S-LWS has an asymptotic distribution independent of the initial robust estimation, is asymptotically efficient under normality, and can be free of auxiliary tuning parameters. Most importantly, the asymptotic distribution independent of the initial robust estimate guarantees that correct inference is easily possible irrespective of the properties of the initial estimator and that the quality of 2S-LWS estimation is not affected by the initial estimator. Consequently, the initial estimator can be chosen to be as robust as possible without concerns about its other qualities and fine-tuning its parameters. On the other hand, the 2S-LWS procedure relies on a data-dependent choice of weights such as those proposed in [6] for REWLS and in [4] for 2S-LWS. Whereas the former weights exhibit a low bias in the presence of outliers, the latter weights lead to a high relative efficiency in small and large samples, but suffer from a sizeable bias in the presence of outliers, especially in time series.

In this paper, we propose new weighting schemes for the 2S-LWS method, which should combine benefits of the weights used in [6] and [4]. Specifically, 2S-LWS with the newly proposed weights should exhibit high finite-sample relative efficiency and low bias in the presence of outliers both in the cross-sectional and time-series applications without rejecting observations. We derive here the robust and asymptotic properties of the proposed methods and document their behavior by simulations.

The rest of this paper is organized as follows. The LWS and 2S-LWS estimators are defined in Section 2. Next, the robust and asymptotic properties of 2S-LWS are studied in Sections 3 and 4, respectively. The finite-sample properties of the proposed method are evaluated and compared with the existing methods using Monte Carlo experiments in Section 5.

## 2. LEAST WEIGHTED SQUARES AND EFFICIENT ROBUST ESTIMATION

Let us consider the linear regression model ( $i = 1, \dots, n$ )

$$(2.1) \quad y_i = x_i^\top \beta^0 + \varepsilon_i,$$

where  $y_i \in \mathbb{R}$  and  $x_i \in \mathbb{R}^p$  represent the response and explanatory variables and  $\beta^0 \in \mathbb{R}^p$  is underlying value of  $p$  unknown regression parameters. [9] proposed to robustly estimate this model by the least trimmed squares (LTS) estimator,

$$(2.2) \quad \hat{\beta}_n^{(\text{LTS})} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{[\lambda n]} r_{[i]}^2(\beta),$$

where  $r_{[i]}^2(\beta)$  represents the  $i$ th order statistics of squared regression residuals  $r_1^2(\beta), \dots, r_n^2(\beta)$  and  $r_i(\beta) = y_i - x_i^\top \beta$ . The trimming constant  $\lambda \in \langle 1/2, 1 \rangle$  determines the robust properties of LTS since definition (2.2) implies that  $n - [\lambda n]$  observations with the largest residuals do not directly affect the estimator.

To improve upon LTS, the author of [14] studied a weighted form of LTS, the least weighted squares (LWS), which can be defined by

$$(2.3) \quad \begin{aligned} \hat{\beta}_n^{(\text{LWS})} &= \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n w\left(\frac{2i-1}{2n}\right) r_{[i]}^2(\beta) \\ &= \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n w\left[G_n\{r_i^2(\beta)\} - \frac{1}{2n}\right] r_i^2(\beta), \end{aligned}$$

where  $w: \langle 0, 1 \rangle \rightarrow \mathbb{R}_0^+$  is a weight function and  $G_n$  denotes the empirical distribution function of squared residuals  $r_i^2(\beta)$ . Note that both LS and LTS are special cases of (2.3) for  $w(t) = 1$  and  $w(t) = I(t \leq c)$ , respectively, for  $t \in \langle 0, 1 \rangle$ . The LWS estimator in the cross-sectional linear regression was extensively studied in [7], where its robust properties, asymptotic normality, and the optimal choice of weight function  $w$  are derived provided that the distribution function of the error term  $\varepsilon_i$  in (2.1) is known. However, to achieve the best robust properties (maximal breakdown point defined later in Section 3), one has to trim almost half of observations and to set  $w(t) = 0$  for  $t > 0.5$ . This results in a low relative efficiency, which is as low as 7% for Gaussian data, and therefore, LWS cannot combine a high breakdown point and a good performance in terms of the estimators' variance.

To combine robustness and high relative efficiency, in [4] we proposed to use data-dependent weights within the LWS estimator. Let us now assume that  $\hat{\beta}_n^0$  and  $\hat{\sigma}_n^0$  are the initial robust estimates of the regression parameters  $\beta^0$  and residual variance

$\sigma^2 = \text{var}(\varepsilon_i)$ . Given model (2.1), the corresponding initial regression residuals are  $e_i^0 = r_i(\hat{\beta}_n^0) = y_i - x_i^\top \hat{\beta}_n^0$ ,  $i = 1, \dots, n$ . For the  $i$ th order residual statistics  $r_{[i]}^2(\hat{\beta}_n^0)$  we can define a weight  $w_i = \hat{w}_n\{(2i-1)/(2n)\}$ , where  $\hat{w}_n$  is a weight function that can generally depend on  $\hat{\beta}_n^0$ ,  $\hat{\sigma}_n^0$  and  $e_i^0$ , but that is assumed to converge to a piecewise continuous function  $w: \langle 0, 1 \rangle \rightarrow \mathbb{R}_0^+$ ,  $\hat{w}_n(t) \rightarrow w(t)$  as  $n \rightarrow \infty$  for all  $t \in \langle 0, 1 \rangle$ . The two-step least weighted squares (2S-LWS) estimator is then defined as

$$(2.4) \quad \hat{\beta}^{(2\text{S-LWS})} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \hat{w}_n \left[ G_n\{r_i^2(\beta)\} - \frac{1}{2n} \right] r_i^2(\beta).$$

A specific feature of the proposed estimator is that the weights modify and apply to the values of the empirical distribution function of the squared regression residuals. Moreover, it might seem that there is asymptotically no difference between 2S-LWS using weights  $\hat{w}_n \rightarrow w$  and LWS using weights  $w$ . The crucial distinction however lies in the fact that the 2S-LWS weight function  $\hat{w}_n$  can converge to an unknown function  $w$  (e.g., depending on the unknown distribution or density functions of  $\varepsilon_i$ ), whereas LWS can be applied only if the weight function  $w$  is known.

Two important examples of data-dependent weight functions used in [4] are the hard-rejection weights adapted from [6] and the everywhere positive quantile-based weights. In both cases, weighting is based on comparison of the empirical distribution function  $G_n^0$  of squared residuals  $r_i^2(\hat{\beta}_n^0)$  and the  $\chi_1^2$  distribution function  $F_\chi$ , which describes the behavior of  $\varepsilon_i^2$  under the assumption of normally distributed errors,  $\varepsilon_i \sim N(0, 1)$ . The hard-rejection weights can be defined by the weight function

$$(2.5) \quad \hat{w}_n^R(t) = I(t < 1 - d_n),$$

where  $d_n = \sup_{t \geq c} \max\{0, \hat{\sigma}_n^0 F_\chi(t) - G_n^0(t)\}$ , the cut-off point  $c$  equals 2.5, and  $\hat{\sigma}_n^0 = [1.4826 \cdot \text{MAD}_{i=1, \dots, n} r_i(\hat{\beta}_n^0)]^2$ . The corresponding procedure, which is equivalent to LTS with a data-dependent choice of trimming  $h_n = [d_n n]$ , is denoted 2S-LWS-R. The everywhere-positive quantile-based weights, which aim to make weighted residuals normally distributed at  $\hat{\beta}_n^0$ , can be defined by

$$(2.6) \quad \hat{w}_n^Q(t) = \hat{\sigma}_n^0 \frac{F_\chi^{-1}(\max\{t, b_n\})}{(G_n^0)^{-1}(\max\{t, b_n\})},$$

where  $(G_n^0)^{-1}$  is the empirical quantile function of  $r_i^2(\hat{\beta}_n^0)$  and  $b_n = \min\{m/n: r_{[m]}^2(\hat{\beta}_n^0) > 0\}$  is used to avoid dividing by zero. The estimator is referred to as 2S-LWS-Q.

Using the hard-rejection weights, which are either zero or one, provides a robust estimator which eliminates extreme inefficiency of LWS for Gaussian data, while fully

preserving the robustness of LWS (if LWS is used as the initial estimator  $\hat{\beta}_n^0$ ). On the other hand, the quantile-based weights employ the information about the whole distribution of regression residuals and thus achieve high relative efficiency even in very small samples, see [4]. Our experiments however show that 2S-LWS-Q exhibits a sizeable bias if there are many outliers in a leverage position, and even more so, if the method is applied in time series containing outliers. To combine advantages of both weighting schemes, we propose here several alternative weighting schemes. Let us first mention that extending  $\hat{w}_n^Q$  by a simple downweighting of observations in leverage position (i.e., observations far from the rest of data in the space of covariates) is not possible without loss of efficiency and we thus do not pursue this strategy.

A straightforward combination of the benefits of the two weight functions  $\hat{w}_n^R$  and  $\hat{w}_n^Q$  could be achieved by using the quantile-based weights for most data points and by fully trimming only the observations with squared residuals highly improbable under Gaussian errors,  $r_i^2(\hat{\beta}_n^0)/\hat{\sigma}_n^0 \geq F_\chi^{-1}(q)$ , where  $0.9999 \leq q < 1$  for instance. The quantile  $q$  can be chosen very close to 1 because 2S-LWS-Q deals well with many outliers, at least if they are not in an extreme or leverage position (cf. [4]). Such a combination of hard-rejection and quantile-based weights can be defined as

$$(2.7) \quad \hat{w}_n^{RQ}(t) = \hat{\sigma}_n^0 \frac{F_\chi^{-1}(\max\{t, b_n\})}{(G_n^0)^{-1}(\max\{t, b_n\})} I(t < 1 - d_n),$$

where  $d_n = \sup_{t \geq c} \max\{0, \hat{\sigma}_n^0 F_\chi(t) - G_n^0(t)\}$  and  $c = F_\chi^{-1}(q)$ . The corresponding LWS estimator with weights  $\hat{w}_n^{RQ}$  is denoted 2S-LWS-RQ( $1 - q$ ).

This straightforward choice of weights however eliminates one of important advantages of the quantile-based weights. Because some observations have zero weight, some regression coefficients could possibly not be identified if there are many categorical explanatory variables. Therefore, it might be preferable to simply downweight observations with extreme residuals more than by  $\hat{w}_n^Q$ . Specifically, we propose the power transformation

$$(2.8) \quad \hat{w}_n^{PQ}(t) = \frac{F_\chi^{-1}(\max\{t, b_n\})}{r[(G_n^0)^{-1}(\max\{t, b_n\})/\hat{\sigma}_n^0]},$$

where  $r(t) = \exp[\ln t \cdot \max\{1, 1 + K d_n[t/F_\chi^{-1}(q) - 1]\}]$ ,  $K \geq 1$  and  $1 > q \geq 0.99$ . In other words, the function  $r$  is linear in its argument (standardized residual) unless the residual is unusually large. Additionally, the power  $1 + K d_n[t/F_\chi^{-1}(q) - 1]$  is proportional to the estimated amount of outlying observations  $d_n$ . The corresponding LWS estimator with weights  $\hat{w}_n^{PQ}$  is referred to as 2S-LWS-PQ( $1 - q$ ). Note that the cut-off quantile  $q$  can be smaller here than in (2.7) because residuals above the cut-off

point do not get zero weights; actually their weights will be close to the original  $\hat{w}_n^Q$  unless they are very large.

### 3. FUNDAMENTAL PROPERTIES OF 2S-LWS

The four weight functions presented in Section 2 describe four weighted two-step robust estimators based on an initial (high-breakdown point) estimator and LWS. In this section we derive asymptotic equivalence of the objective functions of 2S-LWS and LS for Gaussian data and the robust properties of the proposed 2S-LWS methods. The asymptotic distribution of 2S-LWS is studied later in Section 4.

One of reasons motivating 2S-LWS was the lack of efficiency of many highly robust estimators in models with Gaussian errors. To document how 2S-LWS improves upon this, we show now that the proposed weight functions  $\hat{w}_n^R, \hat{w}_n^Q, \hat{w}_n^{RQ}$  and  $\hat{w}_n^{PQ}$  uniformly converge to a constant function on  $(0, 1)$  as  $n \rightarrow \infty$  if  $\varepsilon_i \sim N(0, \sigma^2)$  in (2.1). Hence for normal data, the objective function of 2S-LWS becomes asymptotically identical to the LS criterion. (Note that the following theorem holds also under more general Assumption A introduced later in Section 4.)

**Theorem 3.1.** *Assume that  $(x_i, \varepsilon_i)_{i=1}^n$  in (2.1) forms a sequence of independent and identically distributed random vectors, that  $\varepsilon_i \sim N(0, \sigma^2)$ , and that the initial estimators  $\hat{\beta}_n^0$  of regression parameters  $\beta^0$  and  $\hat{\sigma}_n^0$  of residual variance  $\sigma^2 = \text{var}(\varepsilon_i)$  are  $n^{1/4}$ -consistent,  $\|\hat{\beta}_n^0 - \beta^0\| = \mathcal{O}_p(n^{-1/4})$  and  $|\hat{\sigma}_n^0 - \sigma^2| = \mathcal{O}_p(n^{-1/4})$  as  $n \rightarrow \infty$ . Then for any  $0 < a < b < 1$  and  $k \in \{R, Q, RQ, PQ\}$ , we have*

$$\sup_{t \in (a, b)} |\hat{w}_n^k(t) - 1| = \mathcal{O}_p(n^{-1/2}).$$

**Proof.** The result is a direct consequence of the inequality  $|\hat{w}_n^1(t)\hat{w}_n^2(t) - w^1(t)w^2(t)| \leq |\hat{w}_n^1(t) - w^1(t)||\hat{w}_n^2(t)| + |w^1(t)||\hat{w}_n^2(t) - w^2(t)|$  and Theorem 3.1 and Lemma 4.3 in [4] because  $\hat{w}^2(t)$  can be in all cases chosen so that  $\hat{w}^2(t) \leq 1$  for  $t \in (0, 1)$ ; for example,  $\hat{w}_n^1 = \hat{w}_n^Q$  and  $\hat{w}_n^2 = \hat{w}_n^R$  in the case of  $\hat{w}_n^{RQ}$ .  $\square$

Another feature of the proposed 2S-LWS estimators is that they either trim only a (small) adaptively chosen proportion of observations (2S-LWS-R[Q]), or alternatively, do not trim observations from its objective function at all, just downweight them (2S-LWS-[P]Q). We prove now that this feature does not eliminate or diminish the robust properties of an initial estimator. To formulate this result, we have to introduce a formal definition of the breakdown point. For the sake of simplicity, we consider independent and identically distributed observations  $(y_i, x_i)_{i=1}^n$  (the breakdown point under dependence is generally model-specific; see [5]). The finite-sample breakdown

point of a linear-regression estimator  $\hat{\beta}_n = T(Z_n)$  or variance estimator  $\hat{\sigma}_n = T(Z_n)$ ,  $Z_n = (y_i, x_i)_{i=1}^n$ , can be then defined as

$$\varepsilon_n^*(T) = \frac{1}{n} \min_{m \geq 0} \left\{ m : \max_{I_m = \{i_1, \dots, i_m\}} \sup_{\tilde{Z}_{n, I_m}} \max\{\|T(\tilde{Z}_{n, I_m})\|, \|T(\tilde{Z}_{n, I_m})\|^{-1}\} = \infty \right\},$$

where modified samples  $\tilde{Z}_{n, I_m}$  are created from the original data  $Z_n$  by replacing observations with indices  $I_m$  by arbitrary values, see [10].

Now, we derive the breakdown point of the proposed 2S-LWS variants.

**Theorem 3.2.** *Let  $(y_i, x_i)_{i=1}^n$  be a sequence of independent and identically distributed random vectors which are almost surely in a general position for  $n > p$ . Further, let  $\varepsilon_{\beta_n}^{0*}$  and  $\varepsilon_{\sigma_n}^{0*}$  be the finite-sample breakdown points of initial estimators,  $\hat{\beta}_n^0$  and  $\hat{\sigma}_n^0$  of the regression parameters and residual variance, respectively. Then the finite-sample breakdown points of the 2S-LWS-R, 2S-LWS-Q, 2S-LWS-RQ and 2S-LWS-PQ estimators are larger than or equal to  $\min\{\varepsilon_{\beta_n}^{0*}, \varepsilon_{\sigma_n}^{0*}, \{[(n+1)/2] - (p+1)\}/n\}$ .*

**P r o o f.** The claim of the theorem is a consequence of Theorem 4.2 in [4] because  $\hat{w}_n^k(t) \leq \hat{w}_n^Q(t)$  or  $\hat{w}_n^k(t) \leq \hat{w}_n^R(t)$  for all  $t \in (0, 1)$  and  $k \in \{R, Q, RQ, PQ\}$ .  $\square$

Theorem 3.2 implies that the breakdown point of the proposed 2S-LWS estimators (asymptotically) equals the minimum of the breakdown points of the initial estimators  $\hat{\beta}_n^0$  and  $\hat{\sigma}_n^0$ . It is thus advisable to use highly robust initial estimators in order to obtain 2S-LWS estimates with a high breakdown point.

#### 4. ASYMPTOTICS OF 2S-LWS

In this section we first introduce necessary notation and assumptions and then derive the asymptotic distribution of 2S-LWS.

The distribution functions of  $\varepsilon_i$  and  $\varepsilon_i^2$  in model (2.1) are referred to as  $F$  and  $G$ , respectively, their density functions are denoted  $f$  and  $g$ , provided that they exist, and the corresponding quantile functions are  $F^{-1}$  and  $G^{-1}$ , respectively. The assumptions necessary to derive the asymptotic normality of LWS concern the random variables  $x_i$  and  $\varepsilon_i$  in model (2.1) and the weight function  $w$ . A detailed discussion of the assumptions can be found in [2].

**A s s u m p t i o n 4.1.** Random vectors  $(x_i)_{i \in \mathbb{N}}$  form a weakly stationary absolutely regular sequence with mixing coefficients  $\beta_m$  satisfying

$$m^{r/(r-2)} (\log m)^{2(r-1)/(r-2)} \beta_m \rightarrow 0 \quad \text{as } m \rightarrow \infty \text{ for some } r > 2$$

and have finite  $r$ th moments. Moreover, let  $E(x_i x_i^\top) = Q$  be a nonsingular matrix and  $n^{-1/4} \max_{i=1, \dots, n} \|x_i\| = \mathcal{O}_p(1)$ .

Further, let  $\{\varepsilon_i\}_{i \in \mathbb{N}}$  be a sequence of symmetrically and identically distributed random variables with finite second moments,  $E(\varepsilon_i) = 0$  and  $\text{var}(\varepsilon_i) = \sigma^2$ , and additionally, let  $\varepsilon_i$  and  $x_i$  be independent. The distribution function  $F$  of  $\varepsilon_i$  is absolutely continuous and its probability density function  $f$  is assumed to be bounded and continuously differentiable.

Finally, let  $w: \langle 0, 1 \rangle \rightarrow \mathbb{R}_0^+$  be a non-negative, bounded, and left-continuous function that has a bounded derivative everywhere except for a finite set  $D = \{d_1, \dots, d_J\}$  of points of discontinuity.

The asymptotic results concerning 2S-LWS are presented in the following theorem.

**Theorem 4.2.** *Let Assumption 4.1 hold for a weight function  $w$  and let the two-step least weighted squares estimator  $\hat{\beta}_n^{(2\text{S-LWS})}$  be defined by a bounded weighting function  $\hat{w}_n$  based on initial estimates  $\hat{\beta}_n^0$  and  $\hat{\sigma}_n^0$ . Further, assume that, in probability,  $\hat{w}_n(t) \rightarrow w(t)$  on  $t \in \langle 0, 1 \rangle$  and  $n^{-\alpha} |\hat{w}_n(t) - w(t)| \rightarrow 0$  uniformly on any compact subset of  $(0, 1)$  for some  $\alpha > 0$  as  $n \rightarrow \infty$ . Then the two-step least weighted squares estimator  $\hat{\beta}_n^{(2\text{S-LWS})}$  is  $\sqrt{n}$ -consistent and asymptotically normal,*

$$(4.1) \quad \sqrt{n}(\hat{\beta}_n^{(2\text{S-LWS})} - \beta^0) \xrightarrow{\mathcal{L}} N(0, V_w)$$

as  $n \rightarrow \infty$ , where the asymptotic covariance matrix  $V_w$  equals

$$(4.2) \quad V_w = \frac{Q^{-1} \text{var}[x_1 \varepsilon_1 w\{G(\varepsilon_1^2)\}] Q^{-1}}{[\int \varepsilon w\{G(\varepsilon^2)\} f'(\varepsilon) d\varepsilon]^2}$$

provided that the denominator is positive.

**Proof.** It follows from Theorem 1 and Corollary 5.2 in [4]. □

This theorem, together with Theorem 3.1, implies that the proposed 2S-LWS estimators are asymptotically equivalent to LS for Gaussian data,  $\varepsilon_i \sim N(0, \sigma)$ . For other cases, note that an estimator of the covariance matrix  $V_w$  is proposed in [4].



## 5. FINITE-SAMPLE PROPERTIES

In this section we present a simulation study done to assess the finite-sample behavior of the proposed 2S-LWS estimators and to compare it with existing methods. First, we study all estimators under various distributional assumptions for cross-sectional data (Section 5.1). Later, all experiments are redone in the context of time series (Section 5.2). In these simulation experiments, we compare all four proposed variants of 2S-LWS with LS, the S-estimator set up for the maximum breakdown point (i.e., with Tukey's biweight function and  $c = 1.547$ ; see [10]), and the data-adaptive robust REWLS with hard-rejection weights, see [6] for details. The adaptive estimators use for the initial robust fit the above described S-estimator. The newly proposed 2S-LWS-RQ and 2S-LWS-PQ were computed for  $q \in \{10^{-4}, \dots, 10^{-6}\}$  and  $q \in \{10^{-2}, \dots, 10^{-4}\}$ , respectively, but we present just 2S-LWS-RQ( $10^{-5}$ ) and 2S-LWS-PQ( $10^{-2}$ ) due to strong similarity of results across various choices of  $q$ . All methods are compared by means of their mean squared errors:  $\text{MSE} = S^{-1} \sum_{s=1}^S \|\hat{\beta}_n^{(T,s)} - \beta^0\|^2$ , where  $\hat{\beta}_n^{(T,s)}$ ,  $s = 1, \dots, S$ , are the estimates for  $S$  simulated samples.

### 5.1. Cross-section simulations

We evaluate the performance of all estimation methods for the regression model

$$(5.1) \quad y_i = 0.5 + x_{1i} - 2x_{2i} + \varepsilon_i,$$

where  $x_{1i}, x_{2i} \sim N(0, 1)$  and  $(x_{1i}, x_{2i}, \varepsilon_i)_{i=1}^n$  forms a sequence of independent random vectors. Although we obtained results for sample sizes from  $n = 25$  to 400, we present representative results for  $n = 100$  based on 1000 simulated samples.

The error distributions considered are: NORM—clean Gaussian data  $\varepsilon_i \sim N(0, 1)$ ; STD( $d$ )—data with errors from the Student's distribution with  $d$  degrees of freedom; HET—data exhibiting heteroscedasticity of an unknown form,  $\varepsilon_i \sim N(0, z_i)$ , where  $z_i \sim U(0.25, 4)$  is independent of  $x_{1i}$  and  $x_{2i}$ ; OUT( $a$ )—data contaminated by  $[an]$  (vertical) outliers,  $\varepsilon_i \sim (1 - a)N(0, 1) + aU(-50, 50)$ ; and LOU( $a, l$ )—data contaminated by outliers in a leverage position, where a fraction  $a$  of observations satisfies  $x_{1i}, x_{2i} \sim N(0, 1)$  and  $\varepsilon_i \sim N(0, 1)$  and the complementary fraction  $1 - a$  of observations follows  $x_{1i}, x_{2i} \sim N(l, 1)$  and  $\varepsilon_i \sim U(-50, 50)$ .

Let us now discuss the simulation results for model (5.1) summarized in Tab. 1. The results for models without outliers—NORM, STD, DEXP and HET—are qualitatively the same. Apart from model NORM, where LS is preferable, the classical non-robust LS and high breakdown-point S-estimator are the worst performing methods and all adaptive robust methods outperform them. From the adaptive methods,

the hard-rejecting REWLS and 2S-LWS-R exhibit higher MSEs than 2S-LWS-Q, which only downweights observations with large residuals rather than reject them. The proposed 2S-LWS-RQ and 2S-LWS-PQ are only slightly worse than 2S-LWS-Q and preferable to REWLS in all cases.

Model	Estimation method						
	LS	S	REWLS	2S-LWS			
				R	Q	RQ( $10^{-5}$ )	PQ( $10^{-2}$ )
NORM	0.032	0.122	0.042	0.041	0.034	0.035	0.038
STD(5)	0.052	0.113	0.052	0.051	0.045	0.046	0.047
STD(3)	0.146	0.107	0.059	0.059	0.052	0.053	0.054
DEXP	0.064	0.075	0.049	0.049	0.042	0.042	0.043
HET	0.180	0.183	0.135	0.135	0.113	0.114	0.115
OUT(0.10)	2.463	0.099	0.039	0.038	0.039	0.040	0.040
OUT(0.25)	7.020	0.097	0.046	0.045	0.053	0.052	0.053
OUT(0.40)	9.986	0.101	0.061	0.061	0.067	0.067	0.065
LOUT(0.10,4)	2.711	0.111	0.045	0.044	0.048	0.046	0.045
LOUT(0.25,6)	3.345	0.108	0.068	0.068	0.088	0.081	0.077
LOUT(0.40,8)	4.263	0.151	0.115	0.115	0.222	0.131	0.119

Table 1. Mean squared errors of all methods in various cross-sectional regression models,  $n = 100$ .

Next, for models containing outliers that are not in leverage positions—OUT(0.10), OUT(0.25) and OUT(0.40)—one observes the complete failure of LS in comparison to robust methods. Additionally, the robust S-estimator is stable, but outperformed by all adaptive robust methods. Contrary to models without contamination, the hard-rejecting REWLS and 2S-LWS-R are best because they can completely eliminate outlying observations. The remaining 2S-LWS methods lead to results similar to each other, which are slightly worse than those of REWLS in terms of MSE.

Finally, simulations with data contaminated by outliers in leverage positions—models LOUT(0.10,4), LOUT(0.25,6) and LOUT(0.40,8)—document both the robustness of all estimators except for LS and the increased sensitivity (bias) of 2S-LWS-Q to this type of contamination. The proposed 2S-LWS-RQ and 2S-LWS-PQ are however much closer to the best performing REWLS in this case, especially 2S-LWS-PQ, and they again outperform the S-estimator. Consequently, 2S-LWS-RQ and 2S-LWS-PQ are able to preserve most of the efficiency of 2S-LWS-Q, and at the same time, to reduce the influence of contamination on the estimator.

## 5.2. Time-series simulations

Now we carry out the simulations done in Section 5.1 in the time-series setting. For this purpose, a simple autoregressive process AR(1) of order one is used:

$$(5.2) \quad y_t = 0.5 + 0.8y_{t-1} + \varepsilon_t.$$

Except for the fact that there are no independent explanatory variables  $x_{1i}$ ,  $x_{2i}$ , the data generating processes concerning the error term  $\varepsilon_i$  correspond to those in Section 5.1 with the exception of model OUT( $a$ ). In that case, an AR(1) process  $\{y_t\}_{t=1}^n$  with the errors from the standard normal distribution is generated,  $\varepsilon_i \sim N(0, 1)$ , and subsequently, a uniformly distributed noise  $\varepsilon_i^* \sim U(-25, 25)$  is added to  $[an]$  randomly selected observations.

Model	Estimation method						
	LS	S	REWLS	2S-LWS			
				R	Q	RQ( $10^{-5}$ )	PQ( $10^{-2}$ )
NORM	0.056	0.153	0.068	0.067	0.060	0.061	0.065
STD(5)	0.067	0.107	0.059	0.059	0.055	0.055	0.056
STD(3)	0.060	0.101	0.054	0.053	0.051	0.050	0.051
DEXP	0.067	0.057	0.047	0.047	0.043	0.042	0.043
HET	0.124	0.090	0.091	0.090	0.076	0.075	0.074
OUT(0.05)	3.079	0.160	0.102	0.102	0.421	0.130	0.125
OUT(0.10)	3.569	0.214	0.172	0.174	1.498	0.304	0.271

Table 2. Mean squared errors in various autoregressive models,  $n = 100$ .

Tab. 2 contains all simulation results for model (5.2). Not surprisingly, the results under different distributional models without contamination—NORM, STD, DEXP, and HET—closely correspond to those in Section 5.1. Thus, LS is the best estimator in model NORM, whereas all adaptive robust methods outperform both LS and the S-estimator in all other models. The best performing methods are again those without hard-rejection of observations, 2S-LWS-Q, 2S-LWS-RQ and 2S-LWS-PQ, which have nearly identical MSEs in all models but NORM. In comparison to the cross-sectional experiments, the proposed 2S-LWS-RQ and 2S-LWS-PQ are now sometimes slightly better than the original 2S-LWS-Q.

Next, the results in the presence of outliers—models OUT(0.05) and OUT(0.10)—highlight not only the non-robustness of LS, but also rather high sensitivity of 2S-LWS-Q to outliers in time series. On the other hand, the proposed modifications 2S-LWS-RQ and 2S-LWS-PQ exhibit much smaller MSEs and are closer to the best-performing hard-rejecting REWLS and 2S-LWS-R methods. Hence, the proposed adaptive 2S-LWS methods can combine high relative efficiency in finite samples and low sensitivity to outliers also in time series models.

## 6. CONCLUSION

In this paper new robust estimation methods are introduced, which offer not only robustness in terms of a high breakdown point, but also asymptotic efficiency for Gaussian data and high relative efficiency under many other distributional models in finite samples. In particular, the 2S-LWS-PQ method performs well both in the cross-sectional and time-series setting without fully rejecting any observation, which makes it universally applicable in models with or without many categorical variables, in data with or without outliers, and in cross-section or time-series data. Additionally, let us note that the processing time needed to compute 2S-LWS estimates is comparable to the computation time of the initial high breakdown-point estimators such as LWS or S-estimators and thus all mentioned advantages of 2S-LWS-PQ are achieved just in twice the time needed for computing traditional highly robust estimators.

Finally, although the methods are proposed and discussed in the context of (homoscedastic) linear regression, many extensions are straightforward. This does not include only regression under heteroscedasticity, but also instrumental variable estimation proposed for LWS by [15], nonlinear regression using results of [2], or maximum likelihood estimation [3] as long as the response variable is continuous.

## References

- [1] *N. S. Balke, T. B. Fomby*: Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *J. Appl. Econom.* 9 (1994), 181–200.
- [2] *P. Čížek*: Least trimmed squares in nonlinear regression under dependence. *J. Statist. Plann. Inference* 136 (2006), 3967–3988. zbl
- [3] *P. Čížek*: General trimmed estimation: robust approach to nonlinear and limited dependent variable models. CentER discussion paper. *Econom. Theory*. To appear.
- [4] *P. Čížek*: Efficient robust estimation of regression models. CentER discussion paper Vol. 87. CentER. Tilburg University, Tilburg, 2007.
- [5] *M. G. Genton, A. Lucas*: Comprehensive definitions of breakdown points for independent and dependent observations. *J. R. Stat. Soc., Stat. Methodol. Ser. B* 65 (2003), 81–94. zbl
- [6] *D. Gervini, V. J. Yohai*: A class of robust and fully efficient regression estimators. *Ann. Stat.* 30 (2002), 583–616. zbl
- [7] *L. Mašíček*: Diagnostics and sensitivity of robust models. Unpublished Ph.D. Thesis. Faculty of Mathematics and Physics, Charles University, Prague, 2004.
- [8] *A. Preminger, R. Franck*: Foreign exchange rates: a robust regression approach. *Int. J. Forecasting* 23 (2007), 71–84.
- [9] *P. J. Rousseeuw*: Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, Vol. B (W. Grossman, G. Pflug, I. Vincze, W. Wertz, eds.). Reidel, Dordrecht, 1985, pp. 283–297. zbl
- [10] *P. J. Rousseeuw, A. M. Leroy*: Robust Regression and Outlier Detection. John Wiley & Sons, New York, 1987. zbl
- [11] *S. Sakata, H. White*: High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* 66 (1998), 529–567. zbl

- [12] *J. R. W. Temple*: Robustness tests of the augmented Solow model. *J. Appl. Econometrics* *13* (1998), 361–375.
- [13] *D. Van Dijk, P. H. Franses, A. Lucas*: Testing for ARCH in the presence of additive outliers. *J. Appl. Econom.* *14* (1999), 539–562.
- [14] *J. Á. Višek*: The least weighted squares II. Consistency and asymptotic normality. *Bulletin of the Czech Econom. Soc.* *9* (2002), 1–28.
- [15] *J. Á. Višek*: Instrumental weighted variables. *Austr. J. Stat.* *35* (2006), 379–387.
- [16] *J. Woo*: Economic, political, and institutional determinants of public deficits. *J. Public Economics* *87* (2003), 387–426.

*Author's address:* *P. Čížek*, Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands, e-mail: `P.Cizek@uvt.nl`.